HW 1

Mathematics 127
Mathematical and Computational Methods in Molecular Biology

Fall 2002
UC Berkeley, CA

Nasser M. Abbasi

Fall 2002 Compiled on August 2, 2022 at 8:06am [public]

Contents

1 Problems **3**

2 Problem 1 **6**

3 Problem 2 **12**

4 Problem 3 **17**

5 Problem 4 **19**

6 Problem 5 **21**

7 Problem 6 **25**

1 Problems

Problem Set 1 (due Thursday September 12)
MATH 127: Mathematical and Computational Methods in Molecular Biology

Please work on the starred problem alone.

Problem 1

Find a sequence of Reidemeister moves to untangle this unknot:

Can you do it without passing through a knot diagram with more than seven crossings?

Can you do it without using all three types of Reidemeister moves?

Problem 2

The goal of this exercise is to introduce you to the NCBI website which houses the GENBANK database (the public genome database). The website is at `http://www.ncbi.nlm.nih.gov/`. We will use this site extensively during the semester.

Go to the website and answer the following questions:

- Find the Ebola genome (Zaire Mayinga strain) and display it in FASTA format. How many occurrences of the string "GATTACA" are there in the genome?
- What is the current best estimate of the size of the mouse genome?
- How many amino acids are there in the human topoisomerase TOP IIIa gene?

Problem 3^* [Brunnian links]

a) Construct a link of four components such that the removal of any component leaves a set of unlinked circles.

b) Construct a link of n components such that the removal of any component leaves a set of unlinked circles.

Problem 4

Consider the helix described by the vector equation

$$\mathbf{r}(t) = acos\omega t\mathbf{i} + asin\omega t\mathbf{j} + b\omega t\mathbf{k},$$

where ω is a positive constant. Prove that the tangent line makes a constant angle with the z-axis and that the cosine of this angle is

$$\frac{b}{\sqrt{a^2+b^2}}.$$

Problem 5

Consider two oriented perpendicular lines in R^3. The first, labeled C_1 consists of the set of points $(0,0,z)$ where $-\infty < z < \infty$. The second, C_2 consists of the set of points $(a,y,0)$ where a is a constant and $-\infty < y < \infty$. Let ϵ, δ be constants and consider the contribution to the linking number of the region where $-\epsilon < z < \epsilon$ and $-\delta < y < \delta$ and $a \to 0$ and show that it is $\frac{1}{2}$. In other words, show that the linking number integral reduces to

$$lim_{a\to 0}\frac{1}{4\pi}\int_{-\epsilon}^{\epsilon}\int_{-\delta}^{\delta}\frac{a}{(a^2+y^2+z^2)^{\frac{3}{2}}}dydz$$

and that it is equal to $\frac{1}{2}$.

Problem 6

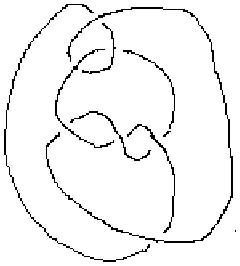
Prove that the following process always will always produce an unknotted diagram: Start drawing. Whenever you encounter a previously drawn line, undercross it. Eventually return to the start.

2 Problem 1

HW1. MATH 127, UC Berkeley.
Due Thursday September 12, 2002.
By Nasser Abbasi (Alone).

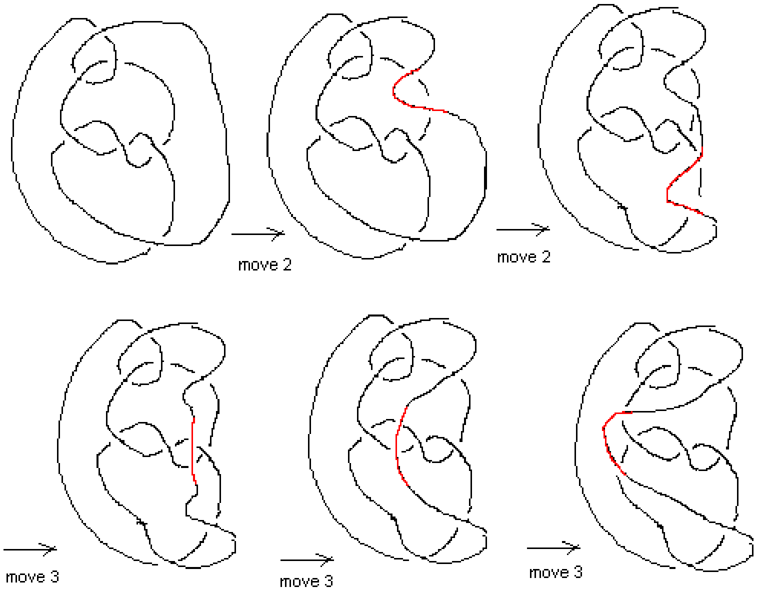
Problem 1

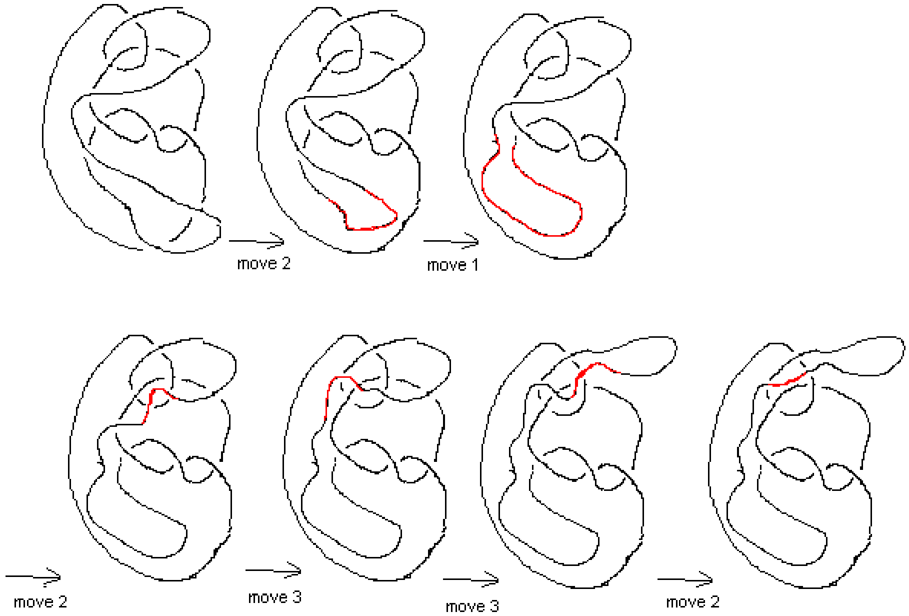
Q: Find a sequence of Reidemeister moves to untangle this unknot:

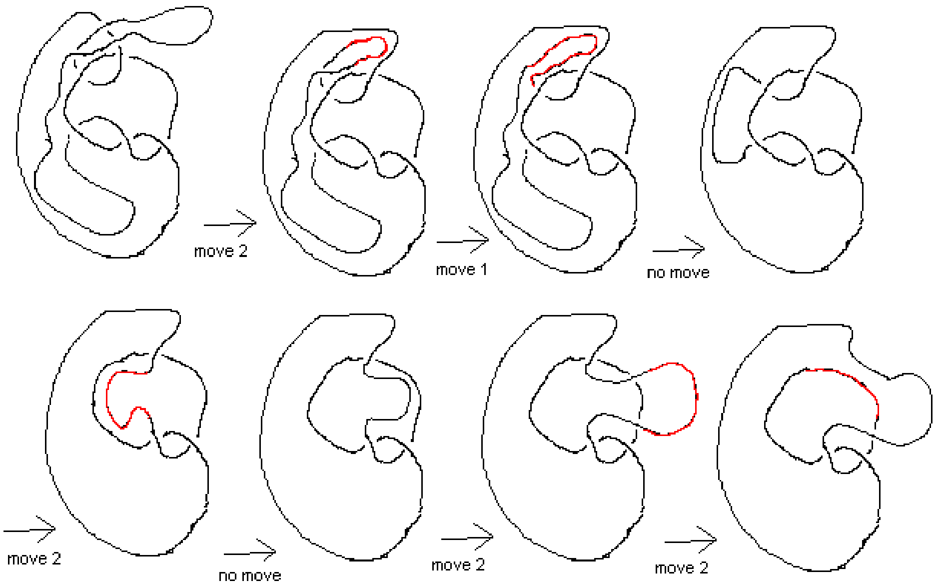


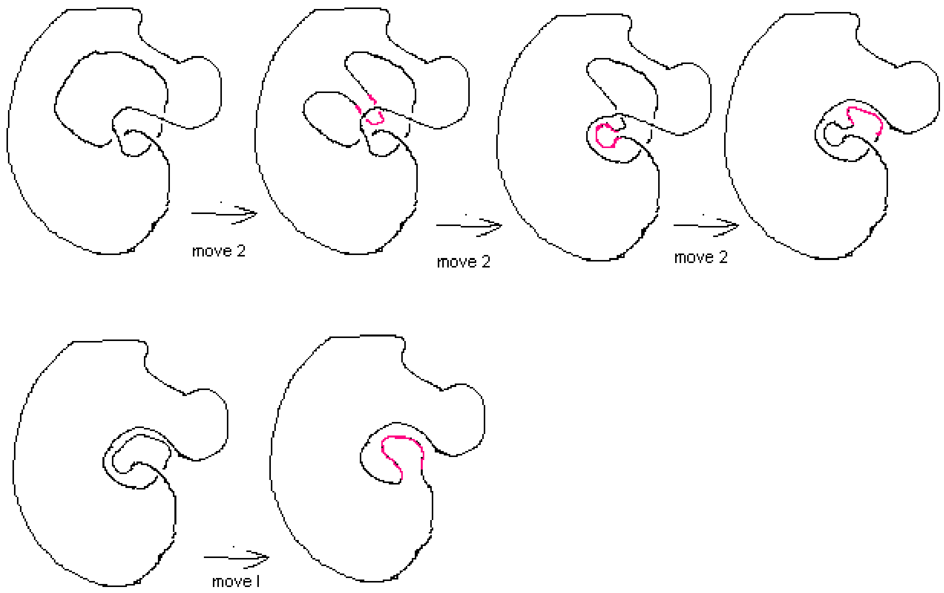
Answer:

The following diagrams show the Reidemeister moves. Diagrams go from left to right, top to bottom. Under each diagram I show the move number (1 which is a twist,2 which is a slide ,or 3 which is a slide over a cross). And I show in RED the part of the diagram that was affected by the Reidemeister move to make it easier to see the move.


move 2
move 2
move 3
move 3
move 3

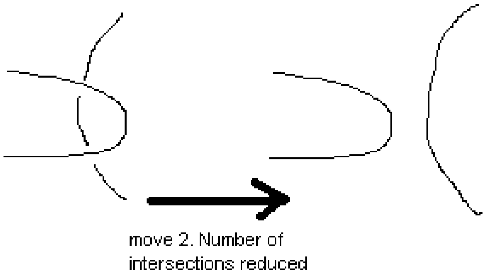

move 2
move 1
move 2
move 3
move 3
move 2

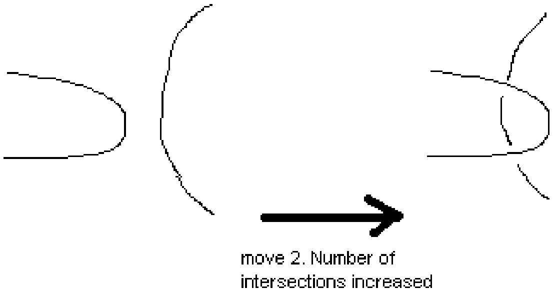

move 2
move 1
no move
move 2
no move
move 2
move 2



Q: Can you do it without passing through a knot diagram with more than 7 crossing?

Answer: NO. Since the initial diagram has 7 crossings, then only applying an untwist move, (move type I), initially will cause the number of crossing to decrease by one. Looking at the initial diagram there is no such initial move.
Hence we must start with a move type 2 or move type 3.
Type 3 move does not change the number of crossing, however in this diagram, there is no such initial move that can be made.
So, what is left is move type 2. This move can either decrease the number of intersections by 2 (call this move 2a type), or increase it by 2 (call this move 2b type) as seen in this diagram:





Looking at the original diagram, there is no possibility of starting with a move 2a type.

This means the only move left to start with is move 2b type, which increases the number of intersections by 2. Hence it is not possible to untangle the unknot without passing through a knot diagram with more than 7 crossings.

Q: Can you do it without using all three types of Reidemeister moves?
A: No. Move type 3 is needed to slide the right most edge to the left over the 3 crossings in the middle of the diagram. If we have started with the left most edge and slide that to the right instead, we still have to use move 3 to pass through the 3 intersections in the middle of the diagram.

3 Problem 2

HW1, problem 2.
MATH 127, UC Berkeley.
By Nasser Abbasi (worked on alone).

Q: Go to http://www.ncbi.nlm.nih.gov and answer the following questions

1. Find the Ebola genome (Zaire Mayinga strain) and display it in FASTA format. How many occurrences of the string "GATTACA" are there in the genome?
2. What is the current best estimate of the size of the mouse genome?
3. How many amino acids are there in the human topoisomerase TOP IIIa gene?

Answer for part 1

!. I went to the above URL. Then selected "Nucleotide" in the search window. Next, typed "Ebola" in the 'for' window. Then clicked GO.

Then a window appeared which listed all GENBANK records for Ebola. I Clicked on the second one in the list (locus AF499101), since that is the strain Zair Mayinga.

A new window appeared showing the GENBANK record for the this virus genome:

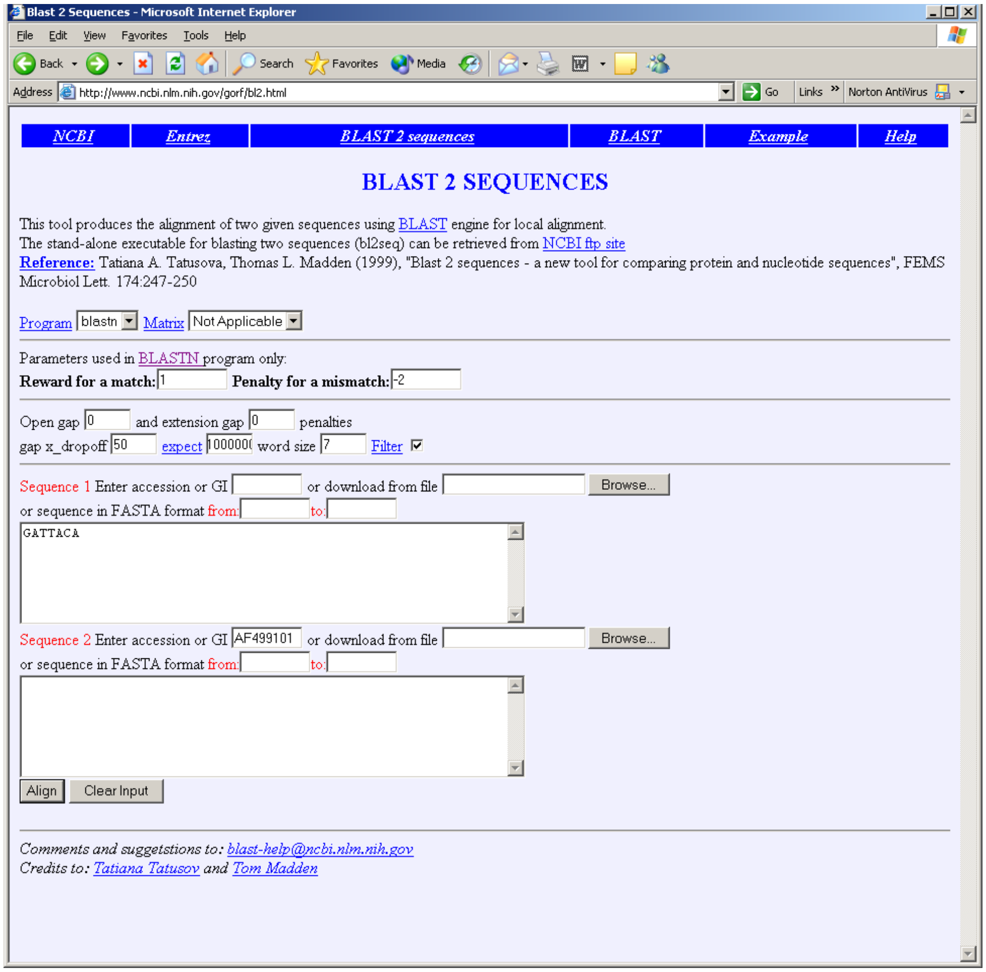
```
LOCUS       AF499101              18960 bp    RNA     linear   VRL 28-AUG-2002
DEFINITION  Zaire Ebola virus strain Mayinga, complete genome.
ACCESSION   AF499101
Etc…
```

In the display option, I selected FASTA, Then clicked on the "TEXT" button. A new window came up showing the geneome in FASAT format.

Showing below few lines of the sequence:

```
>gi|21702647|gb|AF499101.1| Zaire Ebola virus strain Mayinga, complete genome
CGGACACACAAAAAGAAAGAAGAATTTTTAGGATCTTTTGTGTGCGAATAACTATGAGGAAGATTAATAA
TTTTCCTCTCATTGAAATTTATATCGGAATTTAAATTGAAATTGTTACTGTAATCACACCTGGTTTGTTT
CAGAGCCACATCACAAAGATAGAGAACAACCTAGGTCTCCGAAGGGAGCAAGGGCATCAGTGTGCTCAGT
TGAAAATCCCTTGTCAACACCTAGGTCTTATCACATCACAAGTTCCACCTCAGACTCTGCAGGGTGATCC
…..
ATGATAATTAAGACATTGACCACGCTCATCAGAAGGCTCGCCAGAATAAACGTTGCAAAAAGGATTCCTG
GAAAAATGGTCGCACACAAAAATTTAAAAATAAATCTATTTCTTCTTTTTTGTGTGTCCA
```

To answer the question of how many occurances of "GATTACA" in the sequence. I used BLAST 2.0 program. Went to http://www.ncbi.nlm.nih.gov/gorf/bl2.html web page I set the 'EXPECT' value to 1,000,000 and reduced the 'WORD SIZE' from 11 to 7. Typed in the query sequence "GATTACA" in the top search box. Then for the other sequence (the Ebola sequence), entered the ACCESSION value AF499101 which I got from the above query. Next I clicked on the align button:



Address http://www.ncbi.nlm.nih.gov/gorf/bl2.html

BLAST 2 SEQUENCES

This tool produces the alignment of two given sequences using BLAST engine for local alignment.
The stand-alone executable for blasting two sequences (bl2seq) can be retrieved from NCBI ftp site
Reference: Tatiana A. Tatusova, Thomas L. Madden (1999), "Blast 2 sequences - a new tool for comparing protein and nucleotide sequences", FEMS Microbiol Lett. 174:247-250

Program blastn Matrix Not Applicable

Parameters used in BLASTN program only:
Reward for a match: 1 **Penalty for a mismatch:** -2

Open gap 0 and extension gap 0 penalties
gap x_dropoff 50 expect 100000(word size 7 Filter ☑

Sequence 1 Enter accession or GI or download from file
or sequence in FASTA format from: to:

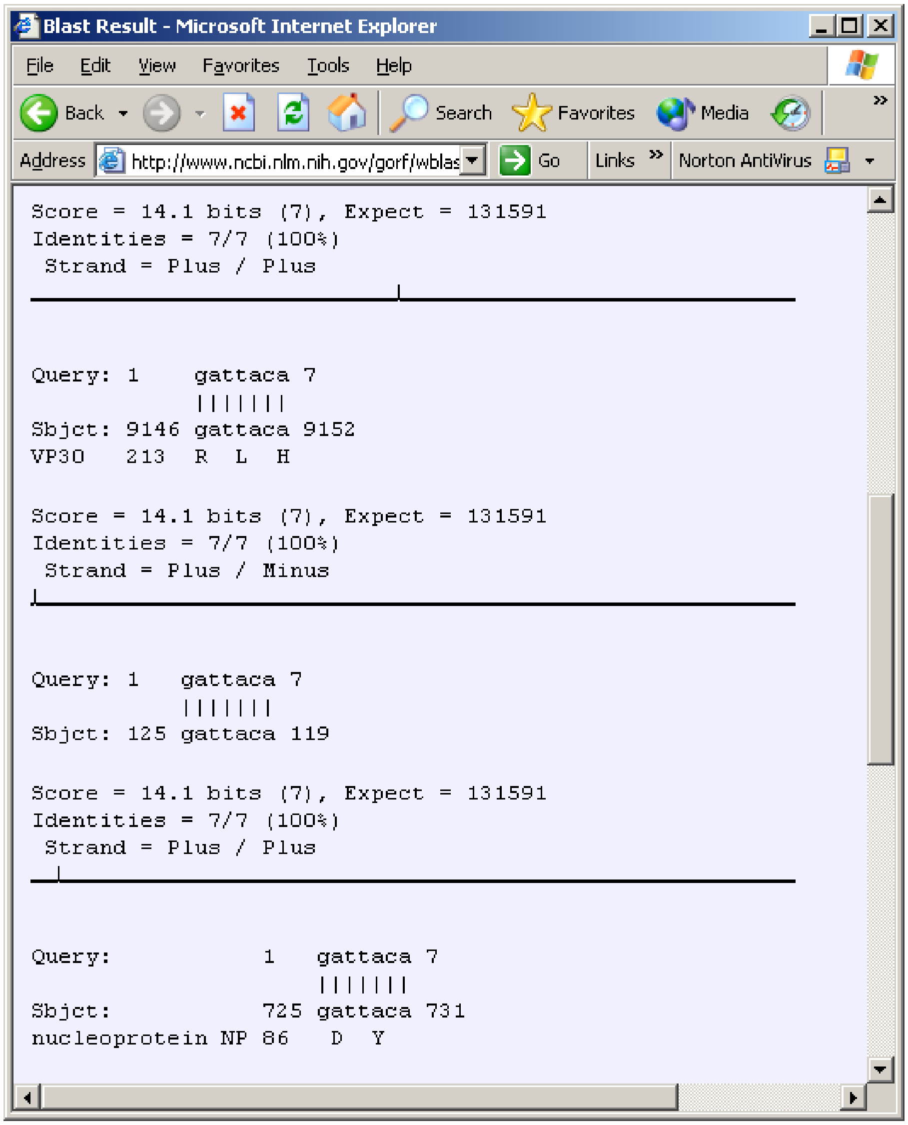
```
GATTACA
```

Sequence 2 Enter accession or GI AF499101 or download from file
or sequence in FASTA format from: to:

Comments and suggetstions to: blast-help@ncbi.nlm.nih.gov
Credits to: Tatiana Tatusov and Tom Madden

The result came back and showed **3 occurances**.

From 9164..9152, 125..119, 725..731



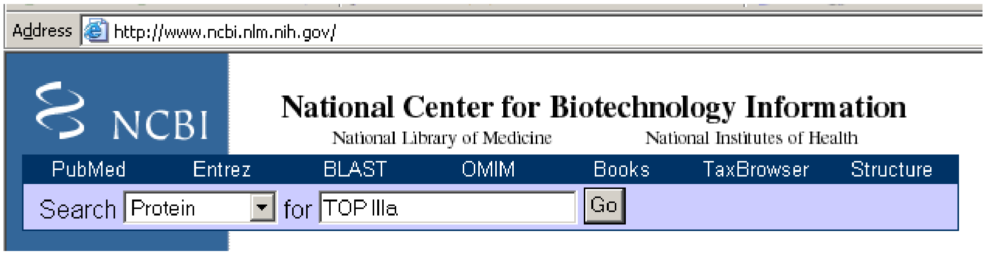
```
Score = 14.1 bits (7), Expect = 131591
Identities = 7/7 (100%)
 Strand = Plus / Plus


Query: 1    gattaca 7
            |||||||
Sbjct: 9146 gattaca 9152
VP30   213  R  L  H

Score = 14.1 bits (7), Expect = 131591
Identities = 7/7 (100%)
 Strand = Plus / Minus


Query: 1   gattaca 7
           |||||||
Sbjct: 125 gattaca 119

Score = 14.1 bits (7), Expect = 131591
Identities = 7/7 (100%)
 Strand = Plus / Plus


Query:            1   gattaca 7
                      |||||||
Sbjct:           725 gattaca 731
nucleoprotein NP 86    D   Y
```

Answer for part 2 question:

To find the best estimate for the size of the mouse genome, went to http://www.ncbi.nlm.nih.gov/genome/seq/MmProgress.shtml
There are 21 chromosoms. The table shows the size of each chromosom in kilo bases (Kb). The total shown is 3,088,000 kb, this is little over 3 billion base pairs. (it is close to the human genome in size).

Answer for part 3 question

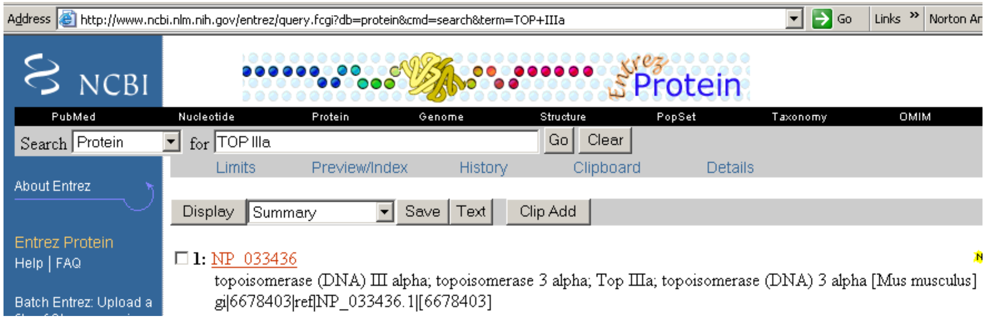
Went to the home page of NCBI again, and selected the Protein database, then typed in 'TOP IIIa' in the 'for' field as shown:



Address http://www.ncbi.nlm.nih.gov/

Search Protein for TOP IIIa

A new page came up with the result:



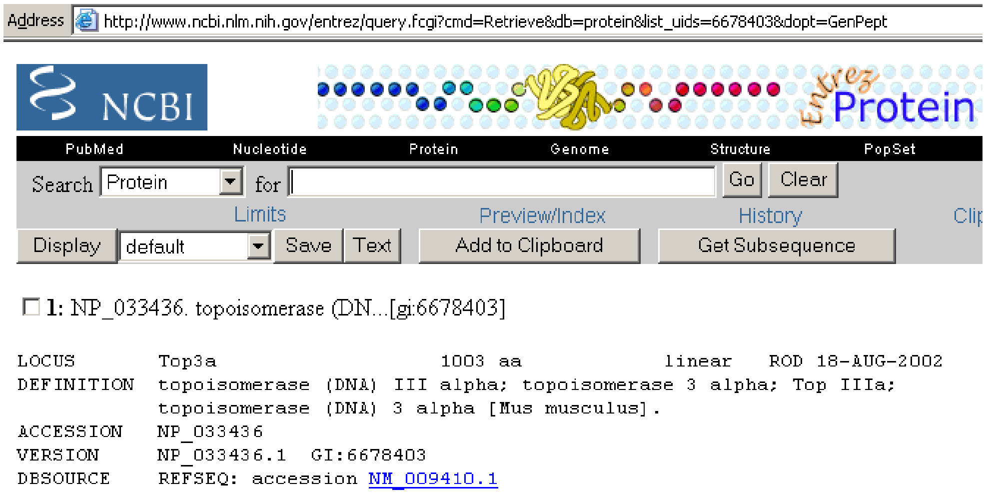
Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=protein&cmd=search&term=TOP+IIIa

Search Protein for TOP IIIa

Display Summary

☐ 1: NP_033436
topoisomerase (DNA) III alpha; topoisomerase 3 alpha; Top IIIa; topoisomerase (DNA) 3 alpha [Mus musculus]
gi|6678403|ref|NP_033436.1|[6678403]

I clicked in the link and got this page:



Address http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=protein&list_uids=6678403&dopt=GenPept

Search Protein for

Display default

☐ 1: NP_033436. topoisomerase (DN...[gi:6678403]

```
LOCUS       Top3a                   1003 aa            linear   ROD 18-AUG-2002
DEFINITION  topoisomerase (DNA) III alpha; topoisomerase 3 alpha; Top IIIa;
            topoisomerase (DNA) 3 alpha [Mus musculus].
ACCESSION   NP_033436
VERSION     NP_033436.1  GI:6678403
DBSOURCE    REFSEQ: accession NM_009410.1
```

The size of the sequence (number of AMINO acids since this is a protein sequence) is **1003**. This can be seen by looking the the features field ‘source’

```
FEATURES             Location/Qualifiers
     source          1..1003
```

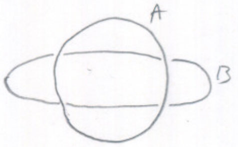
4 Problem 3

HW1 problem 3
Nasser Abbasi

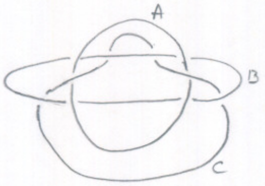
a) First make a 3 component trivial link. then add a link.

This is a 3 component trivial link:

+10
10



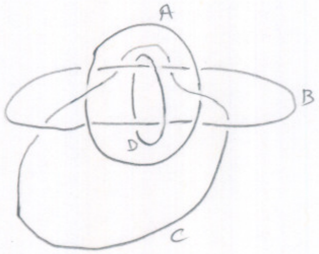
start with 2 links.



add a 3rd component 'c' so that the link remain trivial.

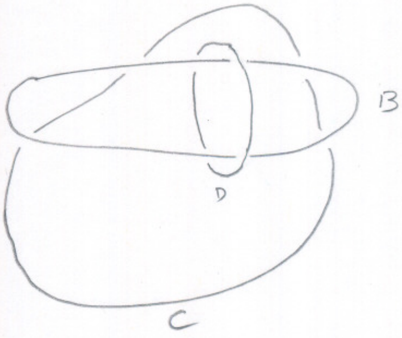
The above is trivial since we can slide 'c' down to cause A circle to slide up freeing all components.

now add a component to cause all 4 to become non-trivial link:



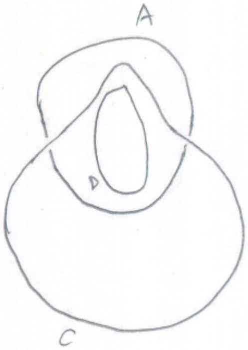
To show that the above is correct, I will remove, one at a time, A, B and C to show that The 3 component link resulting is trivial →

remove A :



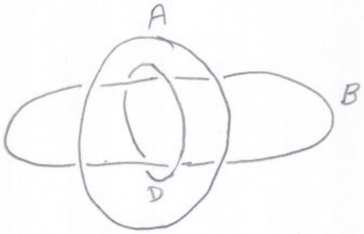
checks out ok.
'C' is separated out
Then 'D' slides out ok.

remove B:



This is clearly a trivial link also.

remove C :



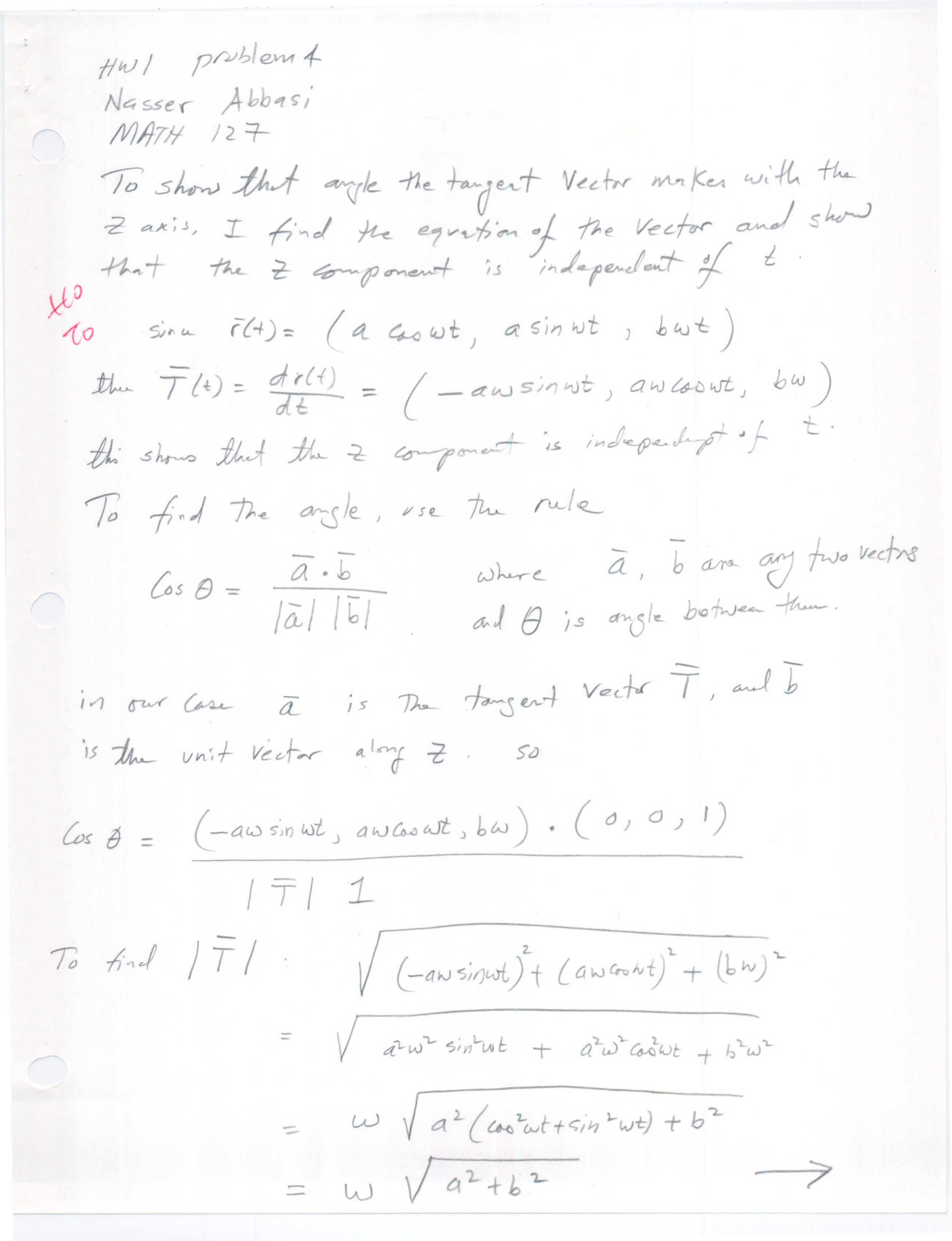
This is also a trivial link.
since 'A' now is free. making
'D' free to slide out.

This completes answer A.

Everyone else looked online.

part (b): unable to do. might be easier to do using Jones Polynomial.
might need to show a series progression somehow.

5 Problem 4



HW1 problem 4
Nasser Abbasi
MATH 127

To show that angle the tangent vector makes with the z axis, I find the equation of the vector and show that the z component is independent of t.

+10
10

since $\bar{r}(t) = (a\cos wt,\ a\sin wt,\ bwt)$

then $\bar{T}(t) = \frac{dr(t)}{dt} = (-aw\sin wt,\ aw\cos wt,\ bw)$

this shows that the z component is indepedept of t.

To find the angle, use the rule

$$\cos\theta = \frac{\bar{a}\cdot\bar{b}}{|\bar{a}|\,|\bar{b}|}$$

where $\bar{a}$, $\bar{b}$ are any two vectrs and θ is angle between them.

in our case $\bar{a}$ is the tangent vector $\bar{T}$, and $\bar{b}$ is the unit vector along z. so

$$\cos\theta = \frac{(-aw\sin wt,\ aw\cos wt,\ bw)\cdot(0,0,1)}{|\bar{T}|\ 1}$$

To find $|\bar{T}|$: $\sqrt{(-aw\sin wt)^2 + (aw\cos wt)^2 + (bw)^2}$

$$= \sqrt{a^2w^2\sin^2 wt + a^2w^2\cos^2 wt + b^2w^2}$$

$$= w\sqrt{a^2(\cos^2 wt + \sin^2 wt) + b^2}$$

$$= w\sqrt{a^2+b^2} \longrightarrow$$

$$\text{So } \cos\Theta = \frac{(-a\omega\sin\omega t,\ a\omega\cos\omega t, b\omega)\cdot(0,0,1)}{\omega\sqrt{a^2+b^2}}$$

$$\cos\Theta = \frac{0+0+b\omega}{\omega\sqrt{a^2+b^2}} = \frac{b}{\sqrt{a^2+b^2}}$$

and angle is constant (independent of t). QED

6 Problem 5

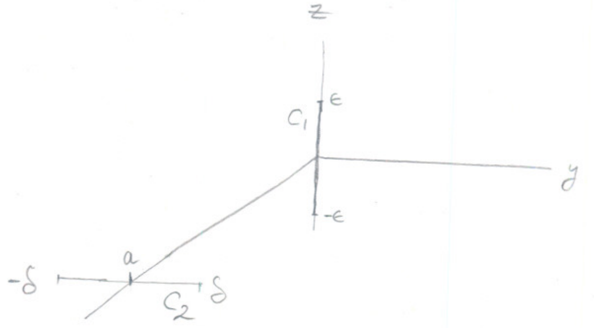
HW 1, problem 5
Nasser Abbasi
MATH 127

a. Consider two oriented perpendicular lines in $\mathbb{R}^3$. the first labeled C_1 consists of the set of points $(0,0,z)$, where $-\infty < z < \infty$. The second C_2 consists of the set of points $(a,y,0)$ where a is a constant and $-\infty < y < \infty$. let ϵ, δ be constants and consider the ~~addition~~ contribution to the linking numbers of the region where $-\epsilon < z < \epsilon$ and $-\delta < y < \delta$ and $a \to 0$ and show that it is $\frac{1}{2}$. In other words, show that the linking number integral reduces to

$$\lim_{a\to 0} \frac{1}{4\pi} \int_{-\epsilon}^{\epsilon} \int_{-\delta}^{\delta} \frac{a}{(a^2+y^2+z^2)^{\frac{3}{2}}}\, dy\, dz.$$

and that it is $= \frac{1}{2}$.

answer



$$LK = \frac{1}{4\pi} \int_{C_2} \int_{C_1} \frac{\bar{e} \cdot (\bar{t}_2 \times \bar{t}_1)}{r^2}\, ds_1 ds_2$$

where $\bar{t}_2$ is unit tangent vector on C_2
and $\bar{t}_1$ is unit tangent vector on C_1
and $\bar{e} = \frac{V_1 - V_2}{\|V_1 - V_2\|}$ where V_1 is eq of point on C_1 and V_2 is eq. of point on C_2.

→

$$\bar{t}_2 = \frac{d}{dy}(a,y,0) = (0,1,0)$$

$$\bar{t}_1 = \frac{d}{dz}(0,0,z) = (0,0,1)$$

so $$\bar{t}_2 \times \bar{t}_1 = \begin{vmatrix} i & 0 & 0 \\ j & 1 & 0 \\ k & 0 & 1 \end{vmatrix} = (1,0,0)$$

$$\bar{e} = \frac{(a,y,0)-(0,0,z)}{\|(a,y,0)-(0,0,z)\|} = \frac{(a,y,-z)}{\sqrt{a^2+y^2+z^2}}$$

so $$LK = \frac{1}{4\pi}\int_{y=-\delta}^{\delta}\int_{z=-\epsilon}^{\epsilon} \overset{\bar{e}}{\frac{(a,y,-z)}{\sqrt{a^2+y^2+z^2}}} \cdot \overset{\bar{t}_1\times\bar{t}_2}{(1,0,0)}\ \frac{1}{r^2}\, dz\, dy.$$

but $r^2 = a^2+y^2+z^2$

so $$LK = \frac{1}{4\pi}\int_{-\delta}^{\delta}\int_{-\epsilon}^{\epsilon} \frac{(a,0,0)}{(a^2+y^2+z^2)^{3/2}}\, dz\, dy. = \frac{1}{4\pi}\int_{-\delta}^{\delta}\int_{-\epsilon}^{\epsilon} \frac{a}{(a^2+y^2+z^2)^{3/2}}\, dz\, dy.$$

to have C_2 and C_1 have a link, 'a' must go to zero else C_2 and C_1 will have link of zero as is.

so $$\boxed{LK = \lim_{a\to 0} \frac{1}{4\pi}\int_{-\epsilon}^{\epsilon}\int_{-\delta}^{\delta} \frac{a}{(a^2+y^2+z^2)^{3/2}}\, dy\, dz}$$

→

To solve the integral. look at $\int_{-\delta}^{\delta} \frac{a}{(a^2+y^2+z^2)^{3/2}}\,dy$.

here, a and Z are constants.

so above integral is $\int a\,(a^2+y^2+z^2)^{-\frac{3}{2}}\,dy$

$$= \frac{4ay}{(4a^2+4z^2)\sqrt{a^2+y^2+z^2}}\Bigg|_{y=-\delta}^{y=\delta} = \frac{2a\delta}{(a^2+z^2)\sqrt{a^2+\delta^2+z^2}}$$

$$\text{so } \frac{1}{4\pi}\int_{-\epsilon}^{\epsilon} \frac{2a\delta}{(a^2+z^2)\sqrt{a^2+\delta^2+z^2}}\,dz = \frac{1}{2\pi}\int_{-\epsilon}^{\epsilon} \frac{a\delta}{(a^2+z^2)\sqrt{a^2+\delta^2+z^2}}\,dz$$

Dr, I used MAPLE to evaluate the above and take the lim a→0, but I get 0 as the final answer not $\frac{1}{2}$. please see attached.

That's okay.

Thanks!!

To compute by hand,

try subs. for polar coords.

$\frac{+10}{10}$

```
> eq1:= (1/(4*Pi))*a*(a^2+y^2+z^2)^(-3/2);
>
```

$$eq1 := \frac{1}{4}\frac{a}{\pi\left(a^2+y^2+z^2\right)^{(3/2)}}$$

```
> eq2:=int(eq1,y=-delta..delta);
```

$$eq2 := \frac{1}{2}\frac{a\,\delta}{\pi\left(a^2+z^2\right)\sqrt{a^2+\delta^2+z^2}}$$

```
> eq3:=int(eq2,z=-epsilon..epsilon);
```

$$\begin{aligned} eq3 := -\frac{1}{4}\delta\sqrt{-a^2}\Bigg(&-\ln\left(-2\,\frac{\delta^2+\varepsilon\sqrt{-a^2}+a^2+\sqrt{\delta^2}\,\sqrt{\varepsilon^2+a^2+\delta^2}}{-\varepsilon+\sqrt{-a^2}}\right) \\ &+\ln\left(2\,\frac{\delta^2-\varepsilon\sqrt{-a^2}+a^2+\sqrt{\delta^2}\,\sqrt{\varepsilon^2+a^2+\delta^2}}{\varepsilon+\sqrt{-a^2}}\right) \\ &+\ln\left(-2\,\frac{\delta^2-\varepsilon\sqrt{-a^2}+a^2+\sqrt{\delta^2}\,\sqrt{\varepsilon^2+a^2+\delta^2}}{\varepsilon+\sqrt{-a^2}}\right) \\ &-\ln\left(2\,\frac{\delta^2+\varepsilon\sqrt{-a^2}+a^2+\sqrt{\delta^2}\,\sqrt{\varepsilon^2+a^2+\delta^2}}{-\varepsilon+\sqrt{-a^2}}\right)\Bigg)\Bigg/\left(a\,\pi\sqrt{\delta^2}\right) \end{aligned}$$

```
> limit(eq3,a=0);
```

$$0$$

Should I have taken $\lim_{\delta \to \pm\infty}$ and $\lim_{\varepsilon \to \pm\infty}$ also ?

7 Problem 6

HW1 problem 6
Nasser Abbasi

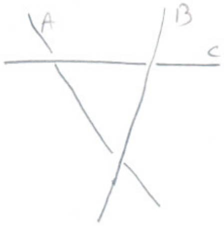
To show this, I need to show that using Reidemeister moves only, the final knot can always be transformed to the unknot (i.e zero crossings).

by definition, we can <u>never get</u> this shape.

+10
10

since c went over and under A which contradicts our method.

we also can <u>never get</u> this shape



since c has gone over A but under B which contradicts our method of building the knot.

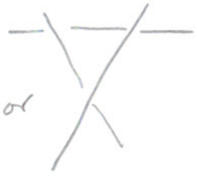
so, the only crossing we can get are

or

or

or

or



or

or

or

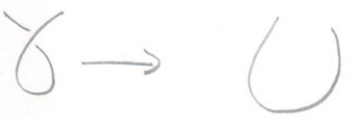
→

Since our goal is to set number of crossing to zero, we use Reim. moves to do this. when we see

Zero crossing

Zero crossing

Zero crossing



Zero crossing.

What is left is Reim. move 3. This move does not change the number of crossing.

Now, I say that the <u>only</u> moves needed to transform This to the unknot are move I and II. (move II is used to slide over intersecting lines one at a time as in

and move I to untwist. so <u>no need to use move III</u>. it will not come up.

This means we can reduce the number of crossings to. Zero.

hence The unknot is produced